

Exercise 1: AdaBoost - Empirical Risk

Let $\hat{f}(\mathbf{x}) = \sum_{m=1}^M \hat{\beta}^{[m]} \hat{b}^{[m]}(\mathbf{x})$ be the scoring function after running AdaBoost for $M \in \mathbb{N}$ iterations. Show that the average empirical risk (on $\mathcal{D}_{\text{train}}$) of the corresponding classifier $\hat{h}(\mathbf{x}) = \text{sign}(\hat{f}(\mathbf{x}))$ is bounded as follows

$$\frac{\mathcal{R}_{\text{emp}}(\hat{h})}{n} = \frac{\sum_{i=1}^n \mathbb{1}_{[\hat{h}(\mathbf{x}^{(i)}) \neq y^{(i)}]}}{n} \leq \prod_{m=1}^M \sqrt{1 - 4(\gamma^{[m]})^2}, \quad (1)$$

where $\gamma^{[m]} = \frac{1}{2} - \text{err}^{[m]}$. For this purpose, proceed as follows:

- (a) Give an interpretation of $\gamma^{[m]}$.
- (b) For any $m = 1, \dots, M$ let $W^{[m]} = \sum_{i=1}^n \tilde{w}^{[m](i)}$ be the total weight in iteration m **before** normalizing the weights. Show that $W^{[m]} = \sqrt{1 - 4(\gamma^{[m]})^2}$.
- (c) Show that

$$w^{[M+1](i)} = \frac{w^{[1](i)} \exp(-y^{(i)} \hat{f}(\mathbf{x}^{(i)}))}{\prod_{m=1}^M W^{[m]}},$$

where $w^{[M+1](i)}$ is the **normalized** weight if we would run AdaBoost for $M + 1$ iterations.

- (d) Argue that $\mathbb{1}_{[\hat{h}(\mathbf{x}) \neq y]} \leq \exp(-y \hat{f}(\mathbf{x}))$ for any $(\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y}$.
- (e) Combine everything to conclude (1).

Hint: Since for any x it holds that $1 + x \leq \exp(x)$, we can infer from (1) that

$$\frac{\mathcal{R}_{\text{emp}}(\hat{h})}{n} \leq \exp \left(-2 \sum_{m=1}^M (\gamma^{[m]})^2 \right) = \exp \left(-2 \sum_{m=1}^M \left(\frac{1}{2} - \text{err}^{[m]} \right)^2 \right),$$

i.e., the average empirical risk is decreasing exponentially in the number of used iterations (provided $\text{err}^{[m]} < 1/2$).